From Attractor Memory to Attention: Dynamical and Synchronization-Based Computation in Neural Networks

Alex Arenas

Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona, Spain.

Corresponding author: alexandre.arenas@urv.cat

Abstract

Associative memory and attention, understood here as context-dependent weighted retrieval, are often discussed as distinct stages in the history of neural computation. However, a dynamical systems perspective reveals a more continuous lineage. Classical Hopfield networks established content-addressable retrieval as collective relaxation toward stored attractors; continuous-time formulations of Hopfield networks made this interpretation explicitly dynamical; and dense and modern Hopfield models subsequently increased storage capacity and revealed a mathematical connection between associative retrieval and the attention mechanism used in transformer architectures. In parallel, a second research tradition replaced static neuronal states with oscillatory degrees of freedom, encoding information in relative phases and retrieving memories through synchronization and phase-locking. This line developed from early phase-oscillator associative memories to experimental and hardware-oriented oscillator computing. Recent work has begun to connect these two traditions, either by deriving attention-like retrieval through softmax-normalized similarity weights as a stationary state of continuous associative memory dynamics or by constructing attention mechanisms through oscillator dynamics and synchronization. This perspective develops these converging trajectories around a common computational question: how can interacting dynamical systems transform an input or query state into the selective retrieval of information from a distributed set of stored or contextual representations? We compare energy minimization, competitive allocation, and synchronization as three dynamical realizations of selective retrieval, examining their mathematical structure, physical realizability, scaling properties, and open problems. This perspective thereby places attention within a broader theory of dynamical associative computation and identifies synchronization-based attention as an emerging, yet experimentally and theoretically immature, direction in physics-native machine learning.

Keywords: associative memory; Hopfield networks; dynamical systems; synchronization; Kuramoto model; oscillator neural networks; attention; transformers; physical neural networks

1. Introduction

Many computational tasks can be formulated as a problem of selective retrieval. A system receives a partial, noisy, or context-dependent state and must identify which information among many stored or currently available representations is relevant. In conventional algorithmic language, retrieval is implemented by an explicit sequence of operations. In dynamical neural systems, however, the computation may instead be carried by the evolution of a high-dimensional interacting system.

This idea is central to the classical Hopfield network. In its original formulation, the physical meaning of content-addressable memory is expressed through a phase-space flow: an incomplete state evolves toward an attractor representing a stored memory [1]. The subsequent continuous

response formulation made the connection to physical relaxation even more direct by replacing binary switching with continuous-time dynamics possessing an energy-like Lyapunov function [2]. Associative retrieval therefore provided an early demonstration that useful computation can emerge from collective dynamics rather than from a prescribed sequence of logical instructions.

A second tradition developed from neural oscillations and synchronization. In this setting, the relevant variables are phases, frequencies, amplitudes, or more general states on compact manifolds. A memory need not be represented by a static state. It can instead be encoded by stable phase relations in a network whose individual units continue to move. Early oscillator associative memory models made this connection explicit by adapting Hebbian ideas to phase couplings and using phase-locking as the retrieval criterion [3, 4]. Later work developed oscillatory neurocomputers, phase-locked loop architectures, experimentally realized oscillator pattern recognizers, and broader synchronization-based computing schemes [5–8].

At first sight, these two traditions appear to concern different problems: attractor neural networks on one side and coupled nonlinear oscillators on the other. The emergence of modern attention mechanisms changes that picture. Dense associative memories and modern Hopfield networks revealed that sufficiently nonlinear associative retrieval can be written in a form closely related, and in a standard formulation equivalent, to the attention update used in transformers [9–12]. Recent work has also shown that softmax-normalized routing can emerge as the fixed point of continuous associative memory dynamics [13]. In parallel, several 2026 works, most of them still preprints, propose attention layers in which synchronization, phase geometry, or oscillator equilibration plays a constitutive role [14–17]. From a complex systems perspective, these models address a common problem: how interactions among many dynamical degrees of freedom generate selective collective states that perform computation.

The central thesis of this perspective is therefore that the two lines are converging realizations of dynamical associative computation. The common question is: How can a network of interacting dynamical degrees of freedom transform a query-like state into selective retrieval from a distributed set of candidate representations? We develop this question through the conceptual map in Fig. 1. The first branch follows energy-based associative memory from classical Hopfield networks to dense and modern Hopfield retrieval and then to attention. The second follows phase-coded associative memory from coupled neural oscillators to synchronization-based computing and, more recently, oscillator-based attention. Competitive routing on a constrained simplex provides an instructive bridge because it generates softmax-like allocation through explicit continuous dynamics rather than by imposing the normalization algebraically [13].

1.1. Scope and status of the literature

This article is a narrative perspective rather than a systematic meta-analysis. It emphasizes work that clarifies conceptual or mathematical connections among associative memory, oscillatory computation, and attention. The synchronization attention literature is especially recent: the most direct proposals appeared in 2026, with one already published in TMLR (Transactions on Machine Learning Research) and several others remaining preprints at the time of writing [14–17]. Claims from this emerging literature should therefore be interpreted according to their current validation and publication status. By contrast, the underlying literatures on Hopfield memories, phase synchronization, oscillator associative memory, and physical oscillator computing span several decades and provide a mature theoretical basis [18–21].

2. Associative memory as attractor dynamics

We begin with associative memory as a dynamical retrieval process, from classical Hopfield attractors to continuous and dense formulations.

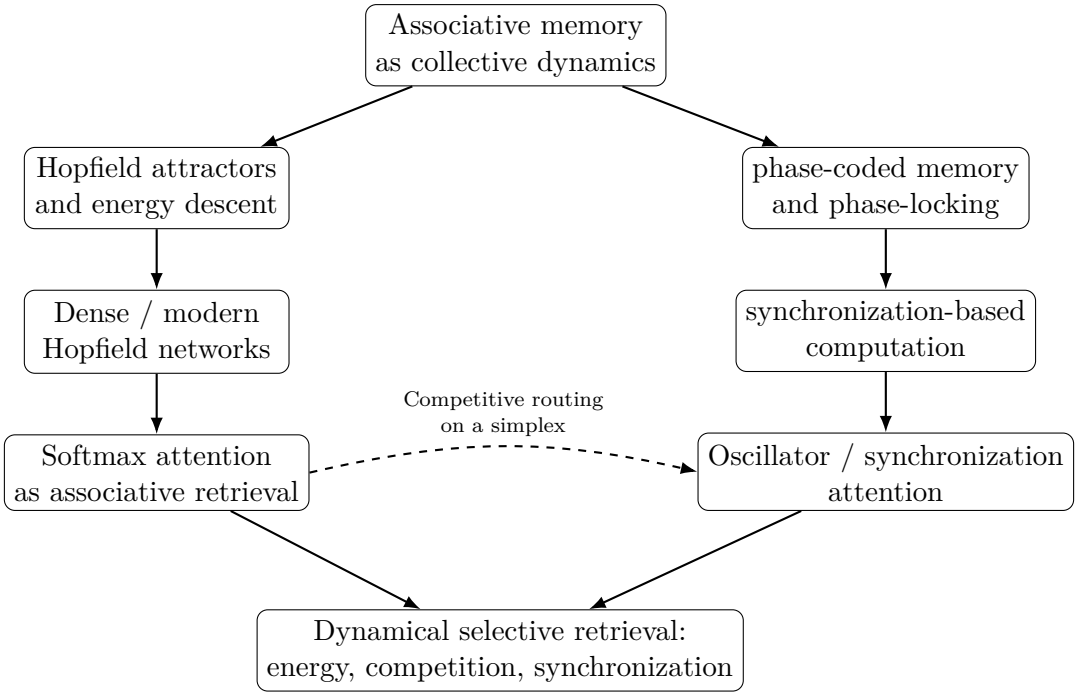

Figure 1: Conceptual organization of the perspective. The diagram is not intended as a strict historical genealogy. It identifies two research traditions that begin from a shared idea—collective dynamical retrieval—and converge on attention-like computation.

2.1. Classical Hopfield networks: memory in the energy landscape

The Hopfield model provides a natural starting point because it formulates retrieval directly as a collective dynamical phenomenon [1]. Consider N binary units $s_i \in \{-1, +1\}$ and a symmetric coupling matrix $W_{ij} = W_{ji}$ with vanishing diagonal. A canonical Hebbian prescription for storing P binary patterns $\boldsymbol{\xi}^\mu$ is

$$W_{ij} = \frac{1}{N} \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu, \qquad i \neq j. \tag{1}$$

Under asynchronous threshold dynamics, the system decreases the energy

$$E(\boldsymbol{s}) = -\frac{1}{2} \sum_{i \neq j} W_{ij} s_i s_j, \tag{2}$$

until it reaches a fixed point. A stored pattern is useful as a memory when it generates a basin of attraction: a sufficiently corrupted version of the pattern is mapped by the dynamics back toward the corresponding attractor.

The computational primitive is therefore not exact lookup but content addressability. The initial condition supplies a query implicitly through overlap with stored memories; the network interactions reshape that partial evidence into a stable collective state. This is a key conceptual point for later comparisons with attention. In both cases, retrieval depends on similarity between the current state and a set of stored or contextual representations. What differs is the mechanism by which similarity is converted into selective weighting.

The classical Hopfield model also exposes the limitations of simple attractor landscapes. Finite storage capacity, crosstalk among patterns, spurious mixture states, and shrinking basins under high load motivate richer energy functions and higher-order interactions [22]. These limitations are not merely historical details: modern associative memories and attention can be understood partly as attempts to make retrieval more selective in large representational spaces.

2.2. Continuous Hopfield dynamics: from state updates to physical flow

The graded response Hopfield model replaces binary neurons by continuous variables while preserving collective retrieval [2]. A common form is

$$\tau_i \dot{x}_i = -x_i + \sum_j W_{ij}\phi(x_j) + I_i, \tag{3}$$

where ϕ is a saturating activation function. For symmetric couplings, suitable choices of dynamics admit a Lyapunov function that decreases along trajectories. Retrieval can then be described with the language of continuous dynamical systems: phase-space flow, basins, fixed points, relaxation times, and stability.

This formulation shifts the interpretation of neural computation. The physical state of the network itself carries out the transformation

$$\boldsymbol{x}(0) \longrightarrow \boldsymbol{x}(t) \longrightarrow \boldsymbol{x}^\star. \tag{4}$$

No external procedure needs to enumerate stored patterns. Instead, the geometry of the flow performs the selection. Continuous-time associative memory thus provides a bridge between statistical mechanical models of neural computation and later efforts to compute directly with physical dynamics.

A recent illustration is the astrocyte gated associative memory model of Vivet and Arenas [13]. That work deliberately begins from a continuous-time Hopfield system and adds a second set of dynamical variables that modulate the contribution of stored patterns. The astrocytic interpretation is motivated by glial modulation of synaptic efficacy, while the gain variables are deliberately coarse grained rather than a model of a specific microscopic pathway. We postpone the attention interpretation to Section 5.3; at this point, the important observation is structural: the classical continuous Hopfield network remains a useful substrate on which additional dynamical degrees of freedom can reshape retrieval.

2.3. Beyond quadratic storage: dense associative memories

The classical quadratic energy is not the only possible associative memory landscape. Dense associative memories replace the low-order interaction structure by stronger nonlinear functions of the overlap with stored patterns [9]. Related mathematical analyses show that sufficiently strong nonlinearities, including exponential interactions, can yield storage capacities that grow far more rapidly with system size than in the classical Hopfield model [10].

Schematically, one may write an energy of the form

$$E(\boldsymbol{x}) = -\sum_\mu F\left(\boldsymbol{\xi}^{\mu\top}\boldsymbol{x}\right), \tag{5}$$

with F controlling how strongly the dynamics discriminate among memories. Increasing the nonlinearity makes retrieval increasingly winner-selective in similarity space. This observation will become important in Section 5, because the exponential weighting underlying modern Hopfield retrieval leads naturally to a softmax form.

Dense associative memories preserve the basic principle of content addressability while making retrieval progressively more selective in similarity space. This sharpening of associative retrieval is one of the two developments that ultimately lead to attention. This route is not exhaustive: Kanerva's sparse distributed memory provides an alternative associative memory tradition, and transformer attention has been related to its retrieval rule under particular conditions [23, 24]. We do not pursue that branch here because the present focus is on dynamical retrieval.

3. From fixed-point memories to phase-locked memories

We next turn to oscillator memories, where information can be encoded in relative phases and retrieved through locking and collective coherence.

3.1. Oscillatory states change the notion of an attractor

A fixed point is not the only way for a dynamical system to represent persistent information. Oscillatory systems can continue to evolve while preserving stable relationships among their components. If

$$z_i(t) = e^{\mathrm{i}\theta_i(t)}, \tag{6}$$

then the absolute phases $\theta_i(t)$ may drift indefinitely while relative phases satisfy

$$\theta_i(t) - \theta_j(t) \longrightarrow \Delta^{\star}_{ij}. \tag{7}$$

A phase-locked configuration is therefore stationary in relational coordinates even though it is not stationary in the laboratory frame.

This change is conceptually important for associative memory. In a Hopfield network, a memory is typically associated with a stable activation vector. In an oscillator memory, information may instead be encoded in a pattern of phase differences, phase clusters, frequency entrainment, or other coherent dynamical relations. Retrieval becomes the formation of the appropriate coherent state.

The Kuramoto model provides the canonical minimal description of synchronization among heterogeneous phase oscillators [18, 25]:

$$\dot{\theta}_i = \omega_i + \frac{\kappa}{N}\sum_{j=1}^{N}\sin(\theta_j - \theta_i). \tag{8}$$

Here κ denotes the Kuramoto coupling strength, distinct from the attention key matrix K introduced later. The natural frequencies ω_i compete with coupling-induced alignment. The order parameter

$$re^{\mathrm{i}\psi} = \frac{1}{N}\sum_{j=1}^{N} e^{\mathrm{i}\theta_j} \tag{9}$$

measures macroscopic phase coherence. The model and its network generalizations became a standard paradigm for synchronization in nonlinear systems [18, 19].

3.2. Phase-locking as associative retrieval

A particularly direct bridge between attractor neural networks and oscillatory dynamics was developed by Arenas and Pérez Vicente [3]. The model considers noisy phase oscillators,

$$\dot{\theta}_i = \omega_i + \gamma_i(t) + \sum_{j=1}^{N} J_{ij}\sin(\theta_j - \theta_i), \tag{10}$$

and stores phase patterns using a Hebbian like coupling rule based on phase differences,

$$J_{ij} = \frac{J}{N}\sum_{\mu=1}^{P}\cos\left(\xi_i^{\mu} - \xi_j^{\mu}\right). \tag{11}$$

The analogy with Eq. (1) is immediate, but the representation is different. A memory is encoded not by a vector of binary activations but by a structured set of phase relations. Retrieval is diagnosed by macroscopic phase correlation with an embedded pattern.

Table 1: Conceptual correspondence between classical attractor memories and oscillator associative memories.

Feature	Hopfield associative memory	Oscillator associative memory
Unit state	Binary spin s_i (continuous activation x_i)	Phase, amplitude, or oscillator state
Stored representation	Activation pattern $\boldsymbol{\xi}^{\mu}$	Phase or synchronization pattern
Interaction	Hebbian coupling W_{ij}	Phase-sensitive coupling J_{ij}
Retrieval	Relaxation to a fixed point	Locking, entrainment, or coherent phase relation
Similarity	State–pattern overlap	Phase overlap or coherence
Typical failure	Crosstalk or spurious attractor	Incoherence, wrong lock, instability, or frequency mismatch
Persistent information	State coordinates	Relative coordinates may carry the memory

The model also highlights a distinction that persists throughout the oscillator-memory literature. Oscillator phases need not be fixed points; the relevant memory variable can be invariant under global rotation. In other words, the attractor lives naturally in a quotient space of relative phases. This makes synchronization a candidate mechanism for storing relational information without requiring the microscopic state to become static.

3.3. Oscillator associative memory after the early phase models

Subsequent work explored the capacity, stability, and architecture of phase-coded associative memories. Aonishi [4] studied a Kuramoto-like oscillator neural network with a standard Hebbian rule and quantified storage and retrieval using statistical mechanical methods. Nishikawa et al. [26] emphasized that naive error free retrieval states in oscillatory memories can be unstable and proposed a modification capable of restoring useful capacity; a more detailed treatment appeared in Nishikawa et al. [27]. These results are important because they prevent a superficial identification of "synchronization" with "successful memory": the topology and stability of coherent states matter.

A complementary line addressed physical realizability. Hoppensteadt and Izhikevich [5] proposed oscillatory neurocomputers in which heterogeneous oscillators, weakly coupled through a common medium, acquire input dependent effective connectivity. Their analysis used Kuramoto-type dynamics and explicitly emphasized associative properties. A later phase-locked loop architecture stored and retrieved complex oscillatory patterns as synchronized states with controlled phase relations [6]. Hölzel and Krischer [7] experimentally demonstrated associative pattern recognition with a network of eight globally coupled electrical oscillators, showing that the phase shifts required for pattern encoding could be controlled in hardware.

The oscillator-memory literature also reveals nontrivial stability issues. In Hebbian Kuramoto networks, attractor geometry can be structurally fragile, and arbitrarily small perturbations may switch the selected output pattern in some regimes [28]. More recent work has revisited oscillatory associative memory away from equilibrium, showing that temporally correlated active noise and nonlinear oscillator interactions can modify retrieval robustness [29]. These studies reinforce an important message for attention oriented applications: if synchronization is to be used computationally, one must analyze not only representational expressivity but also dynamical stability, convergence time, and robustness to heterogeneity.

4. Synchronization as a computational primitive

The oscillator-memory literature suggests a broader interpretation: synchronization can itself be regarded as a primitive computational operation. The transition is conceptual. Phase coherence is first a physical phenomenon, then a representational variable, and finally a mechanism that performs selection or transformation.

4.1. From coherence to information-processing

The Kuramoto order parameter in Eq. (9) summarizes collective alignment. In computation, however, one often needs more structured observables: cluster membership, pairwise phase differences, locked versus drifting populations, or the phase of a selected mode. These observables can implement operations that are naturally relational. For example, synchronized subsets can represent binding or grouping, while competition among synchronization modes can implement selection.

This viewpoint is not restricted to simple phase models. Malagarriga et al. [8] showed that synchronization patterns in networks of neural mass oscillators can perform Boolean like computation and reproduced the mechanism experimentally with coupled Chua circuits. The same physical network could realize different logical mappings under different dynamical inputs. The result illustrates a broader principle: the computational function need not be hardwired solely in static connectivity; it can depend on the dynamical regime selected by the input.

The distinction is particularly relevant to machine learning. Standard artificial networks typically regard internal dynamics as an implementation detail or discretize them into layers. Oscillator computing instead treats phase evolution, entrainment, and relaxation as the operations themselves. Reviews of oscillatory dynamical computing document applications to associative recall, pattern classification, optimization, and neuromorphic processing across electronic, spintronic, mechanical, optical, and other physical substrates [20, 21]. Oscillator based Ising machines provide a related example in which phase dynamics follow a Lyapunov function connected to an Ising Hamiltonian [30]; because their primary task is combinatorial optimization rather than associative retrieval, we do not review them in detail.

4.2. Physical implementation and the promise of native dynamics

Oscillator networks are attractive for unconventional computing because synchronization is a generic physical phenomenon. Proposed or demonstrated substrates include voltage controlled electronic oscillators, phase-locked loops, nonlinear electrical circuits, spin torque and magnetic oscillators, phase transition devices, photonic oscillators, microelectromechanical resonators, and superconducting systems [5, 20, 21]. The specific equations differ across substrates, but weakly coupled systems often admit reduced phase descriptions, creating a common mathematical language.

The potential advantage is not simply analog arithmetic. The stronger claim is that the physical system can solve its own equations: coupling, relaxation, locking, and mode selection occur through the substrate's natural dynamics. This is conceptually aligned with the broader program of physical neural networks, in which trainable computation is delegated directly to optical, mechanical, electronic, or other physical transformations [31, 32].

At the same time, physical dynamics do not make computation free. Coupling networks must be instantiated, states must be initialized, signals must be read out, device mismatch and noise must be tolerated, and a dynamical system may require a non-negligible settling time. These considerations will become central when synchronization is proposed as a replacement for an algebraic operation such as attention.

5. Associative retrieval becomes attention

We now examine how selective associative retrieval connects mathematically to attention and how attention-like weighting can itself emerge from dynamical systems.

5.1. Scaled dot product attention as selective retrieval

Neural attention was introduced for sequence modeling in encoder–decoder architectures [33]; the transformer subsequently made attention the central operation [12]. For a query matrix Q, a key matrix K, and a value matrix V, scaled dot product attention is

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right) V. \tag{12}$$

For a single query $\boldsymbol{q}$, the operation first computes similarities with candidate keys and then converts those similarities into normalized positive weights. The output is a weighted combination of values:

$$\boldsymbol{y} = \sum_\mu a_\mu(\boldsymbol{q})\, \boldsymbol{v}_\mu, \qquad a_\mu = \frac{\exp(\beta\, \boldsymbol{k}_\mu^\top \boldsymbol{q})}{\sum_\nu \exp(\beta\, \boldsymbol{k}_\nu^\top \boldsymbol{q})}. \tag{13}$$

This is already close to the language of associative memory: a query selects among stored representations according to similarity, with the selected content returned through a weighted readout.

Attention and Hopfield retrieval are not identical in every architecture, and the interpretation should not erase distinctions among queries, keys, values, recurrent fixed points, and feedforward transformer layers. Nevertheless, modern Hopfield theory supplies a precise mathematical bridge rather than a merely verbal analogy.

5.2. Modern Hopfield networks and the attention update

The dense memory program of Krotov and Hopfield [9] and the exponential capacity analysis of Demircigil et al. [10] culminated in modern Hopfield formulations with continuous states and exponentially selective retrieval. Ramsauer et al. [11] showed that the corresponding update rule can be written in the same form as transformer attention. In one standard notation, stored patterns are columns of X and the state $\boldsymbol{q}$ is updated as

$$\boldsymbol{q}_{\text{new}} = X\ \text{softmax}\Big(\beta X^\top \boldsymbol{q}\Big)\,. \tag{14}$$

Equation (14) is structurally the same content addressed weighted sum as Eq. (13); the transformer generalizes the construction by allowing distinct learned query, key, and value projections.

This correspondence changes how attention can be situated historically. The progression from classical Hopfield memory through dense associative memory to modern Hopfield retrieval is not merely historical: it establishes a mathematically substantive connection to attention. The selective normalization of attention can be interpreted as an extreme form of similarity dependent associative retrieval, while the values provide a flexible channel separating the object used for matching from the object returned.

The connection also clarifies why the classical Hopfield model is not merely background. The essential operation—using the current state to retrieve relevant stored structure—survives, while the energy landscape, representational dimensionality, and sharpness of retrieval are transformed. Related energy-based transformer constructions make this connection explicit by designing attention layers to minimize an engineered energy function [34]; we regard these as adjacent to, rather than a separate branch of, the present lineage.

5.3. Softmax routing as an emergent fixed point

The model of Vivet and Arenas [13] provides a different route from continuous associative dynamics to attention-like computation. It starts from a continuous Hopfield network and introduces a gain p_μ for each stored pattern. The effective connectivity is

$$W(\boldsymbol{p}) = \frac{P}{N}\sum_{\mu=1}^{P} p_\mu \, \boldsymbol{\xi}^\mu \boldsymbol{\xi}^{\mu\top}, \qquad p_\mu \geq 0, \qquad \sum_\mu p_\mu = 1. \tag{15}$$

Here P denotes the number of stored patterns, as in Eq. (1). The prefactor P/N follows the normalization written here and is not an independent coupling parameter. The gains evolve on the probability simplex under an entropy regularized replicator equation. Writing the pattern-dependent fitness as

$$F_\mu(\boldsymbol{x}, p_\mu) = f_\mu(\boldsymbol{x}) - T \ln p_\mu, \tag{16}$$

the gain dynamics take the replicator form

$$\tau_p \dot{p}_\mu = p_\mu \left[F_\mu - \sum_\nu p_\nu F_\nu \right]. \tag{17}$$

At an interior fixed point, equality of active fitnesses yields

$$p_\mu^\star = \frac{\exp[f_\mu(\boldsymbol{x}^\star)/T]}{\sum_\nu \exp[f_\nu(\boldsymbol{x}^\star)/T]}. \tag{18}$$

softmax-normalized routing therefore appears as a stationary distribution of the gain dynamics. The full neuron–gain system admits a Lyapunov description in the formulation analyzed by Vivet and Arenas [13], so the attention-like routing is embedded in a convergent dynamical system rather than appended as an external normalization.

Two qualifications are important. First, the main model uses a squared overlap fitness, so its default readout is best described as attention-like rather than identified blindly with the exact transformer equation. Second, the same paper notes that a linear overlap variant yields a stationary gain proportional to $\exp(\boldsymbol{\xi}^{\mu\top}\phi(\boldsymbol{x})/T)$, which produces the familiar softmax weighted mixture of stored patterns. The conceptual significance is therefore robust: normalized selective routing can emerge from competitive dynamics on a constrained resource simplex.

This mechanism differs from modern Hopfield attention. In the latter, exponential selectivity is encoded in the associative energy and the update rule. In the astrocyte gated model, normalization is generated by a second dynamical population that competes for finite gain. The two mechanisms thus reach similar computational structures through distinct dynamical architectures. The distinction is computational rather than absolute: the coupled neuron–gain system admits a Lyapunov description, and the stationary softmax allocation has a variational interpretation on the simplex. Competitive routing can therefore also be viewed as a form of constrained energy-based dynamics.

6. Attention through synchronization

The preceding sections establish the two ingredients needed for a synchronization-based interpretation of attention:

(i) attention can be understood as a form of selective associative retrieval; and

(ii) oscillator networks can perform associative retrieval through phase-locking and synchronization.

The resulting question is natural: can the attention operation itself be implemented through oscillator dynamics?

This direction is very recent. The works reviewed below appeared in 2026, and several remain unreviewed preprints. They should be read as alternative formulations of an emerging research program rather than as a settled architecture class.

6.1. Fixed query attention by oscillator equilibration

Pasqualetti and Guo [15] propose attention by synchronization in coupled oscillator networks with the explicit motivation of computing attention on energy-constrained physical substrates. Their construction replaces the softmax arithmetic by equilibration of a gradient flow on a sphere. Learned anchors play a query/key-like role, free oscillator states evolve under Kuramoto–Lohe type dynamics, and the settled geometry is converted into attention weights through a similarity readout.

The high-dimensional setting is important. Classical Kuramoto phases live on $\mathbb{S}^1$, whereas Lohe-type generalizations describe synchronized dynamics on higher dimensional compact spaces [35]. This creates a direct geometric interface with vector embeddings: oscillator states can be unit vectors rather than scalar phases, and cosine-like inner products can serve as readout observables.

The conceptual move is distinct from implementing softmax in analog electronics. The aim is to replace the attention computation by a physical relaxation process whose equilibrium already contains the desired selection information. This places the proposal in the "physics-native" rather than merely "hardware acceleration" category discussed in Section 8.

6.2. Kuramoto attention on a torus

Nunley [16] takes a different route. Hidden coordinates are represented as angles on a torus, and the attention weighted update is constructed so that it coincides with a Kuramoto coupling term. In this setting, the attention matrix acts as an adaptive, content dependent coupling kernel and the value update pulls phases toward a weighted circular mean.

The interpretation reverses the logic of fixed-query oscillator equilibration. Instead of deriving an attention readout from the final state of a physical relaxation, one starts from a trainable attention-like layer and makes the state space and update law explicitly synchronization geometric. The resulting model therefore retains softmax weighting while embedding the value dynamics on a phase manifold.

This distinction is useful for taxonomy. "Oscillator attention" can mean at least two things:

(a) oscillator dynamics replace part of the attention arithmetic; or

(b) oscillator geometry constrains, or interprets, a learned attention layer.

Conflating these categories would obscure their computational and hardware implications.

6.3. Selective synchronization and locking based sparsity

Hays [14] proposes Selective Synchronization Attention, in which tokens are endowed with learnable phases and natural frequencies and pairwise attention is associated with a frequency-dependent phase-locking condition. The attractive conceptual feature is the use of the synchronization threshold as a gating mechanism: oscillator pairs outside the locking regime receive vanishing interaction in the proposed closed form operator.

This idea connects attention sparsity to a dynamical criterion rather than an externally imposed mask. It also revives a central feature of the classical synchronization literature: heterogeneity in natural frequencies is not a nuisance alone; it controls which units can entrain. Because the work is currently a preprint, the scalability and comparative performance claims

require independent validation. Nevertheless, the architecture highlights a general design principle likely to recur in this field: dynamical phase boundaries can act as computational gates. The same threshold also creates a learning question. The learned frequencies and coupling parameters vary smoothly, but setting interactions to zero outside the locking region makes the closed form operator piecewise differentiable: the support of the effective interaction graph changes when a pair crosses the locking boundary. Whether training can exploit these moving boundaries as useful gates without introducing optimization instability remains open.

6.4. Frustration as structured non-synchrony

Perfect synchronization is not always computationally informative. Nunley [17] develops this observation explicitly by constructing attention around frustrated synchronization. The model introduces phase frustration, signed harmonics, and delayed couplings related to Kuramoto–Sakaguchi and Daido type dynamics. The computation is thereby placed in structured departures from complete agreement rather than in convergence to a single synchronized phase.

This is conceptually important beyond the specific architecture. Associative retrieval does not necessarily require maximal coherence. Useful representations may instead correspond to controlled phase offsets, clusters, metastable states, or competing coherent modes. From the perspective developed here, the move from synchronization to structured synchronization parallels the move from a single global Hopfield minimum to a rich energy landscape: computation depends on the geometry of the available collective states, not merely on whether "order" is present. In this sense, frustration can enrich the repertoire of distinguishable collective states rather than merely obstruct synchronization. The analogy is structural rather than one of storage capacity, because the mechanisms generating these state spaces differ from those of dense associative memories.

6.5. A taxonomy of synchronization-based attention

The early literature can be organized by asking six questions:

1. State space: Are token states Euclidean vectors, phases on $\mathbb{S}^1$, points on a torus, or states on a higher dimensional sphere?

2. Similarity to coupling map: Does query–key similarity determine coupling strength, natural frequency compatibility, an anchor field, or a frustration parameter?

3. Dynamical operation: Is attention obtained from equilibrium, one step phase evolution, a phase-locking threshold, or a frustrated coherent state?

4. Readout: Are values combined by cosine similarity, circular means, explicit softmax weights, or another resonant/geometric observable?

5. Physical claim: Is the oscillator formulation primarily interpretive, an inductive bias for software models, or a blueprint for direct physical implementation?

6. Value pathway: Are retrieved values represented independently of the variables used for matching, or does the oscillator state itself carry the returned content?

The value distinction is architecturally important. Selective Synchronization Attention retains a separate learned value projection and uses synchronization to determine the aggregation weights [14], whereas Kuramoto attention uses the phase states themselves as values, making the value update exactly a synchronization step [16]. The latter gives a tighter dynamical interpretation but couples matching geometry to the content being updated. Whether that coupling is a useful inductive bias or a limitation remains open.

Table 2: Three dynamical realizations of selective associative retrieval. The comparison emphasizes computational organization rather than mutually exclusive mathematical classes.

Feature	**Energy-based** retrieval	**Competitive routing**	**Synchronization-based** retrieval
Representative state	Neural or embedding state $\boldsymbol{x}$	State $\boldsymbol{x}$ with gains $\boldsymbol{p}$	Phase or oscillator state $\boldsymbol{\theta}, \boldsymbol{z}$
Driving principle	Descent in an energy landscape	replicator-like competition under a normalization constraint	coupling-induced phase or frequency coherence
Selection variable	Basin, overlap, or exponential energy contribution	Simplex weight p_μ	Locking, phase offset, cluster coherence, or equilibrium geometry
Stationary object	fixed point or metastable minimum	Joint fixed point $(\boldsymbol{x}^\star, \boldsymbol{p}^\star)$	Locked state, synchronized manifold, or time-averaged coherent state
Connection to attention	Softmax attention is recovered for modern Hopfield retrieval under appropriate exponential interactions and parameterization	softmax-normalized gains emerge at stationary points of simplex dynamics; the exact form depends on the fitness	Model dependent: oscillator dynamics may generate weights, gate interactions, or define state updates; some formulations retain explicit softmax
Key physical issue	High-dimensional energy descent	Competitive normalization of gains	Coupling topology, frequency control, settling time, and readout

7. A unifying dynamical perspective

The literature reviewed above suggests three useful dynamical realizations of selective retrieval: associative energy descent, competitive allocation on a constrained simplex, and synchronization. These categories describe computational organization rather than mutually exclusive mathematical classes. Competitive allocation can itself admit a Lyapunov or variational formulation, while some synchronization dynamics are also gradient like. What distinguishes the three routes is therefore primarily the state space, constraints, and mechanism by which similarity or compatibility is converted into selection. Despite these differences, they share a common computational structure: the current state is compared with stored or contextual representations, this similarity or compatibility shapes the collective dynamics, and the resulting dynamics selectively determine the retrieved representation.

The strength of these correspondences is not identical. Modern Hopfield theory provides an explicit algebraic equivalence under specified assumptions, whereas competitive and synchronization-based constructions range from exact recovery of particular weighting rules to alternative dynamical realizations of the broader selective retrieval operation.

7.1. What is common?

All three realizations transform a distributed similarity signal into selective retrieval. In a Hopfield like energy model, memories that overlap strongly with the current state contribute more strongly to the descent direction. In competitive gain dynamics, patterns with greater fitness acquire a larger share of a conserved routing resource. In oscillator networks, states with stronger or more compatible coupling are more likely to lock or dominate the coherent mean field.

This common structure suggests that attention should not be identified exclusively with

softmax. Softmax is one particularly successful map from similarity to normalized selection. The broader computational primitive is context-dependent weighted retrieval. Dynamical systems offer multiple ways to realize such weighting.

7.2. What is fundamentally different?

The differences are equally important. Energy-based models typically assume a scalar potential or Lyapunov function, whereas synchronization systems may admit rotating states, limit cycles, cluster states, or non-gradient dynamics. Competitive routing introduces an explicit normalization manifold, the simplex, whereas oscillator normalization may be geometric, for example through unit norm phase or sphere constraints. Finally, the time variable has different computational meanings. In a transformer, a layer is ordinarily evaluated as an algebraic map. In a physical dynamical system, output quality depends on how long the system is allowed to evolve and on whether the target state is reached robustly.

These differences imply distinct failure modes. Energy-based retrieval can be trapped in undesirable minima. Competitive routing can become overly diffuse or overly winner-take-all depending on regularization and timescale ratios. Oscillator systems can fail to lock, synchronize to the wrong mode, become sensitive to parameter disorder, or require long transients near dynamical boundaries. Normalization itself is not a necessary ingredient of selective retrieval. The simplex constraint normalizes allocation weights, whereas unit norm or phase constraints normalize the state geometry; these serve different mathematical purposes. More generally, selection requires sufficient contrast among candidate representations, but that contrast may also arise through thresholds, winner selection, attractor basin geometry, or dynamical instability without an explicit sum to one constraint.

7.3. The role of time

One of the deepest distinctions in this literature is between attention as an instantaneous algebraic map, and attention as the asymptotic or finite time state of a dynamical system.

The latter raises questions absent from standard algorithmic complexity alone. If a physical network computes by relaxation, then latency depends on the convergence rate. Critical slowing down near bifurcations may trade selectivity for speed. Noise can either destroy a coherent state or, in some nonequilibrium memories, reshape effective retrieval landscapes [29]. A fair theory of dynamical attention therefore requires a joint accounting of operation count, physical time, energy dissipation, precision, and readout overhead.

8. Physical attention: acceleration versus native dynamics

The physical computing motivation should distinguish two strategies that are often grouped together.

8.1. Hardware acceleration of conventional attention

One strategy keeps the mathematical attention operator essentially intact and accelerates its dominant operations on unconventional hardware. Analog in-memory computing is a prominent example. Leroux et al. [36] implement the dot products and activation stages of an attention mechanism using charge-based gain cell arrays and mixed analog–digital circuitry, targeting the latency and energy cost of attention during language model inference.

This is a hardware–algorithm co-design problem: the substrate approximates or modifies standard matrix operations while preserving the general transformer architecture. The physics provides efficient multiplication, storage, or activation, but the desired computation remains recognizably attention.

8.2. Physics-native attention

Here we use *physics-native* to denote a stronger strategy in which the required computation corresponds directly to a natural dynamical process, rather than merely accelerating the same algebraic operation. Oscillator equilibration, phase-locking, and synchronization are examples. The substrate is not merely asked to approximate matrix arithmetic; its intrinsic equations are used as the algorithm.

This distinction matters for evaluating claims of efficiency. If coupling weights must still be computed and communicated digitally, or if readout requires expensive global normalization, the physical advantage may move rather than disappear. Conversely, if a device naturally forms the relevant coherent state in parallel and exposes an inexpensive observable, the computational primitive may be genuinely different from a digital implementation.

The broader physical neural network literature makes the same point. Training and deploying unconventional physical systems requires co-design between learning rules, device physics, calibration, readout, and noise models [31, 32]. Synchronization-based attention should therefore be evaluated as a full physical information-processing pipeline, not only as a differential equation with an attractive interpretation.

9. Open problems

The convergence of associative memory theory, oscillator computing, and attention opens a number of concrete research problems.

9.1. Retrieval capacity and dynamical expressivity

Classical and modern associative memories have well developed notions of storage capacity, basin size, and retrieval error. Oscillator memories have analogous but not identical quantities involving phase overlap, locking stability, and coherent state capacity [4, 26]. Attention models are usually evaluated by downstream task performance rather than by a formal retrieval capacity theory. A useful synthesis would define capacity notions that apply across energy-based, competitive, and synchronization-based retrieval.

A related question concerns the role of the underlying dynamics. Gradient systems offer strong convergence tools, while oscillatory systems derive representational power precisely from the possibility of persistent motion. The field needs a clearer account of when convergence is necessary and when finite time or recurrent dynamics are computationally useful. In particular, structured phase differences, cluster states, and frustration may be more expressive than global synchronization, but they complicate stability analysis.

9.2. Learning under physical and dynamical constraints

Most machine learning architectures assume that weights can be tuned independently with high precision. Physical oscillator arrays rarely provide such freedom. Coupling signs, symmetry, delays, and frequency ranges may be constrained by the substrate [20, 21]. Learning algorithms should therefore optimize within physically realizable coupling families rather than train an unconstrained model and map it to hardware afterward.

Oscillator based attention also raises a distinct question about training. Model parameters must often be learned through dynamical evolution, equilibrium states, or synchronization transitions. Near locking thresholds, small parameter changes can alter which oscillators synchronize, potentially making training unstable or difficult. Slow relaxation may also increase training cost. An important open question is therefore whether oscillator-based attention can be trained reliably at scale, and which learning methods are best suited to these dynamics.

9.3. Finite time, robustness, and readout

If the output is defined by an equilibrium or a locked state, the relevant complexity measure includes settling time. Near synchronization thresholds, relaxation can become slow. Yet those same thresholds may provide useful selectivity or sparsity. Quantifying the trade-off among sharp retrieval, dynamical criticality, latency, and noise tolerance is essential for any claim of hardware advantage.

Oscillator hardware is intrinsically heterogeneous. Natural frequency disorder can be computationally useful because it creates selective entrainment, but it can also destabilize stored phase relations. Classical oscillator-memory work already demonstrates sensitivity and structural instability issues in Hebbian Kuramoto networks [28]. Modern attention oriented systems should therefore report performance under realistic device mismatch, stochastic forcing, drift, and finite measurement precision.

Readout is part of the same physical computation. A physical state is useful only if it can be measured. Phase detection, time averaging, amplitude measurement, and analog to digital conversion can dominate the cost of a physical computing system. Future oscillator attention proposals should therefore specify the observable that represents attention and include its measurement cost explicitly.

9.4. Scaling and experimental validation

The direct synchronization attention literature is currently at an early stage. It remains unclear how oscillator state dimension, number of heads, sequence length, depth, causal masking, positional information, and training stability scale simultaneously. Comparisons should include optimized conventional baselines and hardware aware alternatives rather than only naive softmax implementations. The distinction between asymptotic arithmetic complexity and end-to-end wall clock or energy cost is especially important.

Convincing evidence for a computational advantage will require competitive task quality at meaningful scale together with end-to-end measurements of latency, memory, energy per processed token, settling time, and robustness to noise or device mismatch. For physical implementations, coupling generation, parameter programming, communication, and readout should be included in the accounting rather than treating the oscillator dynamics in isolation.

9.5. Beyond pairwise attention

Oscillator networks naturally support collective order parameters, higher harmonics, frustration, and multi-frequency interactions. These structures may correspond more naturally to higher-order associative retrieval than to pairwise dot product attention. Rather than reproducing the transformer operator exactly, a promising direction may be to identify sequence processing primitives that exploit genuinely collective physics and have no simple pairwise attention analogue.

10. Conclusion

Associative memory, synchronization, and attention can be placed within a common framework of dynamical selective retrieval. Classical Hopfield networks established the core idea: a distributed interacting system can recover relevant information by evolving toward a collective state. Continuous Hopfield models made that computation explicitly physical and continuous in time. Dense and modern associative memories then sharpened similarity dependent retrieval and exposed a direct mathematical connection to transformer attention.

In parallel, oscillator neural networks showed that associative information need not reside in static neuronal states. Memories can be encoded in relative phase and recovered through phase-locking. This line developed into oscillatory neurocomputers, experimentally realized pattern

recognizers, and broader synchronization-based computation. Recent work now brings the two traditions together. Attention-like softmax routing can emerge as a stationary distribution of continuous associative memory dynamics, while new oscillator attention architectures seek to obtain selective contextual retrieval from synchronization, phase geometry, or structured frustration.

The resulting picture is broader than a choice between Hopfield networks and transformers or between digital and analog hardware. It identifies at least three dynamical realizations—energy minimization, competitive allocation, and synchronization—that convert similarity into selective retrieval. Their mathematical differences determine their capacity, stability, speed, robustness, and physical realizability. A mature theory of physics-native attention will need to compare these mechanisms on equal footing, including the cost of dynamics, coupling, initialization, and readout.

The most important open question is therefore not whether oscillators can imitate softmax. It is whether the physics of collective dynamical systems can support forms of selective retrieval that are simultaneously trainable, scalable, robust, and computationally advantageous. Associative-memory theory provides the historical and mathematical language for posing that question; synchronization theory provides a rich family of physical mechanisms with which to answer it.

References

[1] John J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

[2] John J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 81(10): 3088–3092, 1984.

[3] Alex Arenas and Conrad J. Pérez Vicente. Phase locking in a network of neural oscillators. *Europhysics Letters*, 26(2):79–83, 1994.

[4] Toru Aonishi. Phase transitions of an oscillator neural network with a standard hebb learning rule. *Physical Review E*, 58(4):4865–4871, 1998.

[5] Frank C. Hoppensteadt and Eugene M. Izhikevich. Oscillatory neurocomputers with dynamic connectivity. *Physical Review Letters*, 82(14):2983–2986, 1999.

[6] Frank C. Hoppensteadt and Eugene M. Izhikevich. Pattern recognition via synchronization in phase-locked loop neural networks. *IEEE Transactions on Neural Networks*, 11(3): 734–738, 2000.

[7] Robert W. Hölzel and Katharina Krischer. Pattern recognition with simple oscillating circuits. *New Journal of Physics*, 13:073031, 2011.

[8] Daniel Malagarriga, Mariano A. García-Vellisca, Alessandro E. P. Villa, Javier M. Buldú, Jordi García-Ojalvo, and Antonio J. Pons. Synchronization-based computation through networks of coupled oscillators. *Frontiers in Computational Neuroscience*, 9:97, 2015.

[9] Dmitry Krotov and John J. Hopfield. Dense associative memory for pattern recognition. In *Advances in Neural Information Processing Systems*, volume 29, pages 1172–1180, 2016.

[10] Mete Demircigil, Judith Heusel, Matthias Löwe, Sven Upgang, and Franck Vermet. On a model of associative memory with huge storage capacity. *Journal of Statistical Physics*, 168 (2):288–299, 2017.

[11] Hubert Ramsauer, Bernhard Schäfl, Johannes Lehner, Philipp Seidl, Michael Widrich, Lukas Gruber, Markus Holzleitner, Thomas Adler, David Kreil, Michael K. Kopp, Günter Klambauer, Johannes Brandstetter, and Sepp Hochreiter. Hopfield networks is all you need. 2020.

[12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pages 5998–6008, 2017.

[13] Arnau Vivet and Alex Arenas. Emergent self-attention from astrocyte-gated associative memory dynamics. *Physical Review Research*, 8(2):023276, 2026.

[14] Hasi Hays. Selective synchronization attention. *arXiv preprint arXiv:2602.14445*, February 2026.

[15] Fabio Pasqualetti and Taosha Guo. Attention by synchronization in coupled oscillator networks. *Transactions on Machine Learning Research*, September 2026.

[16] Joshua Nunley. Kuramoto attention: Synchronizing self-attention on the torus. *arXiv preprint arXiv:2606.11585*, June 2026.

[17] Joshua Nunley. Attention as frustrated synchronization. *arXiv preprint arXiv:2606.18694*, June 2026.

[18] Juan A. Acebrón, Luis L. Bonilla, Conrad J. Pérez Vicente, Félix Ritort, and Renato Spigler. The kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, 77(1):137–185, 2005.

[19] Alex Arenas, Albert Díaz-Guilera, Jürgen Kurths, Yamir Moreno, and Changsong Zhou. Synchronization in complex networks. *Physics Reports*, 469(3):93–153, 2008.

[20] Arijit Raychowdhury, Abhinav Parihar, Gus Henry Smith, Vijaykrishnan Narayanan, Gyorgy Csaba, Matthew Jerry, Wolfgang Porod, and Suman Datta. Computing with networks of oscillatory dynamical systems. *Proceedings of the IEEE*, 107(1):73–89, 2019.

[21] Aida Todri-Sanial, Corentin Delacour, Madeleine Abernot, and Filip Sabo. Computing with oscillators from theoretical underpinnings to applications and demonstrators. *npj Unconventional Computing*, 1(1):14, 2024.

[22] Daniel J. Amit. *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge University Press, Cambridge, 1989.

[23] Pentti Kanerva. *Sparse Distributed Memory*. MIT Press, Cambridge, MA, 1988.

[24] Trenton Bricken and Cengiz Pehlevan. Attention approximates sparse distributed memory, 2021. arXiv:2111.05498.

[25] Yoshiki Kuramoto. Self-entrainment of a population of coupled non-linear oscillators. In *International Symposium on Mathematical Problems in Theoretical Physics*, volume 39 of *Lecture Notes in Physics*, pages 420–422. Springer, 1975.

[26] Takashi Nishikawa, Ying-Cheng Lai, and Frank C. Hoppensteadt. Capacity of oscillatory associative-memory networks with error-free retrieval. *Physical Review Letters*, 92(10): 108101, 2004.

[27] Takashi Nishikawa, Frank C. Hoppensteadt, and Ying-Cheng Lai. Oscillatory associative memory network with perfect retrieval. *Physica D: Nonlinear Phenomena*, 197(1–2):134–148, 2004.

[28] Robert W. Hölzel and Katharina Krischer. Stability and long term behavior of a hebbian network of kuramoto oscillators. *SIAM Journal on Applied Dynamical Systems*, 14(1): 188–201, 2015.

[29] Matthew Du, Agnish Kumar Behera, and Suriyanarayanan Vaikuntanathan. Active oscillatory associative memory. *The Journal of Chemical Physics*, 160(5):055103, 2024.

[30] Tianshi Wang and Jaijeet Roychowdhury. OIM: Oscillator-based ising machines for solving combinatorial optimisation problems, 2019. arXiv:1903.07163.

[31] Logan G. Wright, Tatsuhiro Onodera, Martin M. Stein, Tianyu Wang, Darren T. Schachter, Zoey Hu, and Peter L. McMahon. Deep physical neural networks trained with backpropagation. *Nature*, 601:549–555, 2022.

[32] Ali Momeni, Babak Rahmani, Benjamin Scellier, Logan G. Wright, Peter L. McMahon, Clara C. Wanjura, Yuhang Li, Anas Skalli, Natalia G. Berloff, Tatsuhiro Onodera, Ilker Oguz, Francesco Morichetti, Philipp del Hougne, Manuel Le Gallo, Abu Sebastian, Azalia Mirhoseini, Cheng Zhang, Danijela Marković, Daniel Brunner, Christophe Moser, Sylvain Gigan, Florian Marquardt, Aydogan Ozcan, Julie Grollier, Andrea J. Liu, Demetri Psaltis, Andrea Alù, and Romain Fleury. Training of physical neural networks. *Nature*, 645:53–61, 2025.

[33] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *International Conference on Learning Representations*, 2015.

[34] Benjamin Hoover, Yuchen Liang, Bao Pham, Rameswar Panda, Hendrik Strobelt, Duen Horng Chau, Mohammed J. Zaki, and Dmitry Krotov. Energy transformer. In *Advances in Neural Information Processing Systems*, volume 36, pages 27532–27559, 2023. doi: 10.52202/075280-1197.

[35] M. A. Lohe. Non-abelian kuramoto models and synchronization. *Journal of Physics A: Mathematical and Theoretical*, 42(39):395101, 2009.

[36] Nathan Leroux, Paul-Philipp Manea, Chirag Sudarshan, Jan Finkbeiner, Sebastian Siegel, John Paul Strachan, and Emre Neftci. Analog in-memory computing attention mechanism for fast and energy-efficient large language models. *Nature Computational Science*, 5(9): 813–824, 2025.